Impact of Commercial Airline Network Evolution on the U.S. Air Transportation System

9th USA/Europe Air Traffic and Management R&D Seminar

June 14th, 2011

Purdue University
Daniel DeLaurentis
Tatsuya Kotegawa

Federal Aviation Administration
Joseph Post
Kimberly Noonan

Contact: ddelauere@purdue.edu
Air Transportation System (ATS) Network Evolution

Between 1990 and 2005:

- 2,500 new links were formed
- 1,700 existing links were removed
- 6,400,000 more operations
- 2,700 links (routes) between nodes (airports)

* Links refer to service routes provided by the airlines with at least 365 flights per year

* New links are counted multiple times if they appear, disappear and reappear again

Annual breakdown of ATS network restructuring

Evolution of Southwest Airlines ('90-'05)

Between 1990 and 2005:

- 2,500 new links were formed
- 1,700 existing links were removed
- 6,400,000 more operations
- 2,700 links (routes) between nodes (airports)

* Links refer to service routes provided by the airlines with at least 365 flights per year

* New links are counted multiple times if they appear, disappear and reappear again
Complexity sources in ATS Network Dynamics

Combination of numerous complex factors result in ATS network restructuring

- Competition between airlines
- Airline influence on airports (hub/focus cities and gates, terminal ownership)
- Fluctuations in demand
- Airport infrastructure
- Aircraft and ATM technologies

Airline specific network restructuring

Number of Air Travel Passengers

- United
- Delta
- Southwest
- American

Number of links added (+) and deleted (-)
Current FAA Traffic Forecasts

FAA needs an understanding of air traffic dynamics to...

- Identify operational shortfalls
- Estimate benefits of future investments and technology
- Determine workforce requirements

... and many more

However, current FAA air traffic forecast methods …

- … only predict traffic growth on existing routes
- … assume no flight routes between cities are added or removed

Research Need: Interject and explore how network evolution impacts result of air traffic forecast
Network Forecast Algorithm

The goal is to construct an algorithm which project where new links are formed and existing links are removed as well as the change in operations due to it, **only using network topology metrics for the predictor variables**

**Pros:**
- Limiting forecast factors only to network topology keeps the required data fields low and data collection is fairly easy
- Aggregate the complex market and business competition behavior into a simple network growth algorithm (from historical data)

**Cons:**
- Geographical/demographical factors are difficult to implement
- Tweaking the network growth algorithm to represent a change in business strategy or other economic scenarios can be difficult
The “Big” Picture

**Inputs**

- Static network topology (based on historical data)
- Future network forecast algorithm
- Restructured Network

**System-wide Evaluation**

- NASPAC (operations)
- Purdue FLEET (environmental)

**Evaluation Criteria**

- Level of congestion / delay
- Emissions

- System-wide evaluation tools are used to evaluate impact of new tech. concepts and policies to US Air Transportation System (ATS) but uses existing or static network structure

- However, system-wide simulations usually span over long periods of time when the airline network can drastically change

- **Objective**: Construct a forecast algorithm which can address network dynamics and implement it into system-wide simulations
Network Topology Metrics (1)

**Link Weight** \( (A_{ij}^w) \):
Strength of the links, or number of operations between node \( i \) and \( j \)

**Node Weight** \( (s_i) \)
Total number of operations at node \( i \)

**Node Degree** \( (k_i) \):
Total number of links at node \( i \)
Network Topology Metrics (2)

**Clustering Coefficient (CC$_i$):**

Nodal value that describes ‘local cohesiveness’ or ‘clustering’

\[
CC_i = \frac{\text{number of closed triplets}}{\text{number of connected triples of vertices}}
\]
Network Topology Metrics (3)

Eigenvector Centrality ($x_i$):

Measure of nodal importance. Google’s page ranking algorithm is a variant of this measure

\[ x_i = \frac{1}{\lambda} \sum_{j \in M(i)} x_j = \frac{1}{\lambda} \sum_{j=1}^{N} A_{i,j} x_j \]

Airports with high EVC may have…

<table>
<thead>
<tr>
<th>Few, but strong connections</th>
<th>Lots of ‘neighbors’</th>
<th>Lots of ‘rich neighbors’</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Diagram" /></td>
<td><img src="image2.png" alt="Diagram" /></td>
<td><img src="image3.png" alt="Diagram" /></td>
</tr>
</tbody>
</table>
Data and Assumptions

Data Source

- Data obtained from BTS Transtats Form 41 database (1990 – 2008)
- Commercial, IFR and domestic flights only
- Number of airports in network: 304
- Number of links: ~2,500

Assumptions

- All airline networks are merged together
- Airports with average annual ops less than 365 and outside the CONUS are omitted
- Links with annual ops less than 365 are omitted
Link Addition Logic Overview

A 3-step procedure:

1. Support Vector Machines (SVM) algorithm filters the number of new link candidates
2. Logistic regression assigns connection probability (fitness value) to each filtered candidate
3. Candidates with highest fitness value are actually connected
Support Vector Machines (SVM)

- Supervised machine learning used for pattern recognition, data classification and regression

SVM example

Data points belonging to group A
Data points belonging to group B
Classification planes produced by SVM
Support Vector Machines (SVM)

- Supervised machine learning used for pattern recognition, data classification and regression
- The forecast algorithm SVM uses eigenvector centrality as a predictor variable to separate unconnected node pairs into the “form connection” and “remain disconnected” groups
- Variable 1: higher evc of the node pair
- Variable 2: lower evc of the node pair
Logistic regression

1. SVM reduce new link candidates from 45,000 to 6,000 with 80-95% accuracy
2. Still need to downsize to ~140 new links per year
3. Assign fitness value to filtered candidates using logistic regression probabilities and connect node pair candidates with top fitness values

\[ P_{\text{connect,ij}} = \frac{1}{1 + \exp(-BX)} \]

1. Predictor variables (\(X\)): clustering coefficient eigenvector centrality, node weight and degree
2. Coefficients (\(B\)): fitted with historical data
Link removal logic

\[ P_{\text{disconnect},ij} = \frac{1}{1 + \exp(0.0019A_{i,j}^w)} \]

- Logistic regression curve that calculates link removal probability is trained
- Predictor variable is solely based on the link weights
- Disconnect probability is used as fitness values (similar to link addition logic)
- Links with highest \( P_{\text{disconnect}} \) are removed in the following timestep
- Higher link weight \( \Rightarrow \) less likelihood of the link to disconnect
Accuracy (% of actual new or removed links found)

Link Addition

- Link addition accuracy remains relatively low due to the large number of node-pair candidates for new connections.
- Annually, there are only ~140 new connections out of 45,000 candidates (>0.3%)!
- Many candidate pairs with similar topology characteristics

Link Removal

- Higher accuracy than link addition, since the candidate pool is much smaller.
- ~90 links are removed out of ~1800 candidates.
- Years with lower link addition accuracy tends to show higher link removal accuracy.
In the study that follows, “Original Schedule” refers to the FAA’s forecast for 3/19/2020 using their standard forecast method. The “Modified Schedule” is generated for the same future day using our network restructuring algorithm.

Both are analyzed in the FAA’s National Airspace System Performance Analysis Capability (NASPAC).
Schedule Modification Results (1)

Change in operations after network restructuring at OEP 35 airports in 2020

*difference in total ops between original and modified schedule is 0.33%
Schedule Modification Results (2)

- As a result of restructuring, the most percentage of new links were added to medium size airports
  - “Modified Schedule” has larger number of airports w/degree between ~20 and ~80
- Distribution of operations (traffic) on routes remained similar (right hand chart)
Schedule Modification Results (3)

- Total delay time was approximately 5% greater (13,000 minutes for one day) in modified schedule.
- Most of the increase seen on the ground; small airborne delay increase due primarily due to arrival queue.
Conclusions and Future Work

Some Concluding Remarks

• Forecast algorithm using network topology parameters shows good accuracy (compared to random draw)
• Extrapolating from historical trends in network restructuring, operations at some large hub airports will decrease but operations at medium size hubs will increase
• Dispersion of traffic from large => medium hubs will increase delay originating from airports (surface, gates and etc.)

Research Questions Remaining:

• How would the inclusion of a “hub fitness” (based on other-than-network parameters, e.g., capacity, demographics) measure change/improve the results?
• Is the network growth logic changing with respect to time (are their discernable dynamics at different epochs)?
• How can we translate new business strategies, policy and regulations into parameters of the forecast algorithms?
Backup Slides
Schedule Modification Results (3)

- NASPAC automatically trims operations when they are scheduled too closely together (due to airport runway capacity)
- Network restructuring has little effect on final trim values
## Accuracy Forecast Algorithm Breakdown

### Link Removal

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Removed Links</td>
<td>119</td>
<td>144</td>
<td>109</td>
<td>94</td>
<td>96</td>
<td>107</td>
<td>108</td>
<td>93</td>
<td>55</td>
<td>48</td>
<td>74</td>
<td>107</td>
<td>59</td>
<td>62</td>
<td>97</td>
<td>132</td>
<td>57</td>
<td>126</td>
<td>94</td>
</tr>
<tr>
<td>Number of Correctly Removed</td>
<td>49</td>
<td>37</td>
<td>43</td>
<td>41</td>
<td>41</td>
<td>48</td>
<td>38</td>
<td>34</td>
<td>24</td>
<td>38</td>
<td>48</td>
<td>18</td>
<td>29</td>
<td>33</td>
<td>38</td>
<td>37</td>
<td>44</td>
<td>38</td>
<td></td>
</tr>
<tr>
<td>Number Forecasted</td>
<td>93</td>
<td>93</td>
<td>93</td>
<td>93</td>
<td>93</td>
<td>93</td>
<td>93</td>
<td>93</td>
<td>93</td>
<td>93</td>
<td>93</td>
<td>93</td>
<td>93</td>
<td>93</td>
<td>93</td>
<td>93</td>
<td>93</td>
<td>93</td>
<td></td>
</tr>
<tr>
<td>Number of Candidates</td>
<td>1601</td>
<td>1574</td>
<td>1531</td>
<td>1574</td>
<td>1833</td>
<td>1853</td>
<td>1676</td>
<td>1687</td>
<td>1738</td>
<td>1877</td>
<td>1898</td>
<td>2051</td>
<td>2146</td>
<td>2222</td>
<td>2299</td>
<td>2281</td>
<td>2395</td>
<td>1854.1</td>
<td></td>
</tr>
</tbody>
</table>

### Link Addition

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Added Links</td>
<td>92</td>
<td>98</td>
<td>112</td>
<td>128</td>
<td>152</td>
<td>127</td>
<td>131</td>
<td>104</td>
<td>106</td>
<td>187</td>
<td>95</td>
<td>260</td>
<td>154</td>
<td>136</td>
<td>174</td>
<td>114</td>
<td>170</td>
<td>182</td>
<td>140</td>
</tr>
<tr>
<td>Number of Correctly Added (EVC Filter Only)</td>
<td>79</td>
<td>84</td>
<td>97</td>
<td>101</td>
<td>131</td>
<td>100</td>
<td>100</td>
<td>85</td>
<td>82</td>
<td>158</td>
<td>84</td>
<td>206</td>
<td>137</td>
<td>125</td>
<td>161</td>
<td>102</td>
<td>148</td>
<td>134</td>
<td>117</td>
</tr>
<tr>
<td>Number Forecasted (EVC Filter only)</td>
<td>6,381</td>
<td>6,415</td>
<td>6,560</td>
<td>6,451</td>
<td>6,507</td>
<td>6,299</td>
<td>6,126</td>
<td>6,141</td>
<td>6,330</td>
<td>6,442</td>
<td>6,486</td>
<td>6,517</td>
<td>6,693</td>
<td>7,382</td>
<td>7,199</td>
<td>7,325</td>
<td>7,074</td>
<td>7,145</td>
<td>6,635</td>
</tr>
<tr>
<td>Number of Correctly Added (EVC Filter + Logistic Regression)</td>
<td>8</td>
<td>16</td>
<td>10</td>
<td>13</td>
<td>15</td>
<td>16</td>
<td>11</td>
<td>13</td>
<td>11</td>
<td>19</td>
<td>16</td>
<td>13</td>
<td>15</td>
<td>13</td>
<td>18</td>
<td>17</td>
<td>11</td>
<td>21</td>
<td>14</td>
</tr>
<tr>
<td>Number Forecasted (EVC Filter + Logistics Regression)</td>
<td>140</td>
<td>140</td>
<td>140</td>
<td>140</td>
<td>140</td>
<td>140</td>
<td>140</td>
<td>140</td>
<td>140</td>
<td>140</td>
<td>140</td>
<td>140</td>
<td>140</td>
<td>140</td>
<td>140</td>
<td>140</td>
<td>140</td>
<td>140</td>
<td></td>
</tr>
<tr>
<td>Number of Candidates</td>
<td>44,455</td>
<td>44,482</td>
<td>44,525</td>
<td>44,519</td>
<td>44,482</td>
<td>44,423</td>
<td>44,403</td>
<td>44,360</td>
<td>44,369</td>
<td>44,318</td>
<td>44,179</td>
<td>44,005</td>
<td>43,910</td>
<td>43,834</td>
<td>43,757</td>
<td>43,775</td>
<td>43,661</td>
<td>44,202</td>
<td></td>
</tr>
</tbody>
</table>
For 2019, the stats for ATL is

- 295 links
- 1,176,100 annual ops
- TAF=>1.9% growth
- Above picture shows a subset of the nodes connected to ATL (#s are link ops)

For 2020:

- Evolve the network to 2020
- For actions involving ATL, 10 links disconnected and 8 new links formed (net decrease of 4 links)
- No weights (ops on links) assigned yet
**STEP 3:**

- Annual ops to new links are added using historical survey (based on link distance)
- Forecasted annual terminal ops @ ATL is 1,196,000 for 2020 (+1.9% from 2019)
- Run Fratar algorithm via 1% error tolerance

**STEP 4:**

- Link weights are adjusted via Fratar so the annual TAF matches terminal ops @ each airport
STEP 5:
Scaling down 2020 =>11/20/20

- Normalize the 2020 network by its total ops
- Calculate total IFR daily ops for 11/20/20 (25,938) and multiply
- Generate schedules based on link weights
- The above results do not reflect actual values for the NASPAC input file since this was an example using a subset of the entire network